

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

RESULT 2


```

[1]
RN      SEQUENCE FROM N.A.
RC      STRAIN=R5;
RX      MEDLINE=2005576; PubMed=10589844;
RA      Teramoto M., Futama H., Harayama S., Watanabe K.;
RT      "Characterization of a high-affinity phenol hydroxylase from Comamonas
RT      testosteroni K5 by gene cloning, and expression in Pseudomonas
RT      aeruginosa PAO1c.";
RL      Mol. Gen. Genet. 262:552-558(1999).
CC      -|- SIMILARITY: CONTAINS A SIGMA-54 FACTOR INTERACTION ATP-BINDING
CC      DOMAIN.
DR      EMBL: AB024741; BAA87867.1; -.
DR      InterPro: IPR003593; AAA_Atpase.
DR      InterPro: IPR002197; HTH_Fis.
DR      InterPro: IPR002078; Sig54_interact.
DR      InterPro: IPR004096; V4R.
DR      Pfam: PF02954; HTH_8.1.
DR      Pfam: PF00158; Sigma54_activat; 1.
DR      Pfam: PF02830; V4R: 1.
DR      PRINTS: PR01590; HTHFIS.
DR      SMART: SM00382; AAA; 1.
DR      TIGRFAMs: TIGR01199; HTH_fis; 1.
DR      PROSITE: PS00675; SIGMA54_INTERACT_1; 1.
DR      PROSITE: PS00676; SIGMA54_INTERACT_2; UNKNOWN_1.
DR      PROSITE: PS0068E; SIGMA54_INTERACT_3; 1.
DR      PROSITE: PS50045; SIGMA54_INTERACT_4; 1.
KW      ATP-binding; DNA-binding; Transcription regulation.
SQ      SEQUENCE 584 AA; 64686 MW; 94AB4D5612513158 CRC64;

Query Match          29.4%; Score 56.5; DB 2; Length 584;
Best Local Similarity 33.3%; Pred. No. 7.5;
Matches 18; Conservative 3; Mismatches 6; Indels 27; Gaps

QY      5 WG-DTLNCWML-----SAFSR-----YARCLAEG---HDGP 31
      || || || || || || || || || || || || || || || ||
Db      155 WGPQDQFSCWMLLGYAGSYSSAFRRPVFFKEMQCSTCGHACHLIEGRFQHEWP 208

RESULT 8
Q8WSN8
ID      Q8WSN8      PRELIMINARY;      PRT;      680 AA.
AC      AC
CD      Q8WSN8;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT      01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE      Hypothetical 77.4 kDa protein.
GN      Y41d4B.26.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC      Rhabditidae; Pelodierinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
[1]
RN      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2.
RX      MEDLINE=99069613; PubMed=9851916;
RA      None;
RT      "Genome sequence of the nematode C. elegans: a platform for
RT      investigating biology. The C. elegans Sequencing Consortium.*;
RT      Science 282:2012-2018(1998).
RT      [2]
RN      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      Gisel C., Lamar B.;
RT      "The sequence of C. elegans cosmid Y41d4B.";
RT      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RT      [3]
RN      SEQUENCE FROM N.A.
RC      STRAIN=BRISTOL N2;
RA      Waterston R.;
RT      "Direct Submission.";
RT      Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RA      EMBL; AC024776; XAL32241.1; -.
DR      InterPro: IPR000536; Hormone_rec_1ig.

```


DR PRINTS; PR00179; LIPOCALIN.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
SQ SEQUENCE 191 AA; 21444 MW; 1424BD9878512F61 CRC64;

Query Match 28.4%; Score 54.5; DB 6; Length 191;
Best Local Similarity 33.3%; Pred. NO. 4.6;
Matches 10; Conservative 6; Mismatches 13; Indels

QY 5 WGD^TLNCWMLSA-FSRYARCLAEFGHDGPTQ 33
||| : | : : ||| | - | - |
Db 112 WGSIYSVVVVVD^TDYKEFALLYSEGAKGPQG 141

RESULT 12

Q9UT16	ID	Q9UT16	PRELIMINARY;	PRT;	287 AA.
Q9UT16	CD	Q9UT16			
Q9UT16	DT	01-MAY-2000	(IREMBLrel. 13, Created)		
Q9UT16	DT	01-MAY-2000	(IREMBLrel. 13, Last sequence update)		
Q9UT16	DT	01-MAY-2000	(IREMBLrel. 13, Last annotation update)		
DE	FE	Yeast atp12 protein precursor homolog.			
DE	FE	SPAC9.12C.			
GN	OS	Schizosaccharomyces pombe (fission yeast).			
OC	CC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	CC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	CC	Schizosaccharomycetes.			
OX	NB	NCBI_TaxID=4896;			
RP	LN	[1]			
RP	LN	SEQUENCE FROM N.A.			
RC	STRAIN=97ZH+;				
RL	RA	Wedler H., Duesterhoeft A., Lyne M.H., Rajandream M.			
RL	RA	Submitted (Oct-1999) to the EMBL/GenBank/DDJB database			
DR	EMBL;	AL121764; CAB57430.1; -			
DQ	SEQUENCE	287 AA; 33149 MW; 211780CDD/B2FFD97 CRC6			

Query Match 28.1%; Score 54; DB 3; Length 287;
Best Local Similarity 47.4%; Pred. No. 8.3;
Matches 9; Conservative 2; Mismatches 8; Indels

```
QY      5 WGDTLNCWMLSAFSRYARC 23
      | :|| | | :|| | .|
Db     198 WLSSLNSWQLAAFERSVSC 216
```

RESULT 13

Q9AAPO	PRELIMINARY;	PRT;	316 AA.
Q9AAPD			
Q9AAPP			
DT	01-JUN-2001 (TrEMBLrel. 17, Created)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Hypothetical protein CC0557.		
GN	CC0557.		
OS	Caulobacter crescentus.		
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;		
CA	Caulobacter.		
OX	NCBI_Taxid=155892;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 19089 / CBI5;		
RX	MEDLINE=21173698; PubMed=11259647;		
RA	Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,		
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,		
RA	Potocka R.I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,		
RA	DeBooy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,		
RA	Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,		
RA	Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,		
RA	Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;		
RL	"Complete genome sequence of <i>Caulobacter crescentus</i> .";		
RT	Proc. Natl Acad Sci U.S.A. 98:4136-4141(2001).		
DR	EMBL; AE005729; AAK22543.1; -.		
DR	TIGR; CC0557; -.		
KW	Hypothetical protein; Complete proteome.		

SQ SEQUENCE 316 AA; 35026 MW; 41C4289216FED963 CRC64;

Query Match 28.1%; Score 54; DB 16; Length 316;
Best Local Similarity 56.2%; Pred. No. 9.2;
Matches 9; Conservative 2; Mismatches 5; Indels

QY 6 GDTLNCWMLSAFSRYA 21
|||:|||||
Db 301 GDILSCW^KLGAVPRYS 316

RESULT 14

Q8R057	PRELIMINARY;	PRT;	179 AA.
ID	Q8R057		
AC	Q8R057;		
DT	01-JUN-2002 (TReMBUrel. 21, Created)		
DT	01-JUN-2002 (TReMBUrel. 21, Last sequence update)		
DT	01-JUN-2002 (TReMBUrel. 21, Last annotation update)		
DE	Similar to aryl-Hydrocarbon interacting protein-like 1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID:10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=EYE;		
RA	Strausberg R.;		
RL	Submitted (APR-2002) to the EMBL/GenBank/DDJB databases.		
DR	EMBL; BC028285; AAH28285.1; -		
SQ	SEQUENCE 179 AA; 20424 MW; 32ED79C343761A10 CRC64;		

Query Match 27.9%; Score 53.5; DB 11; Length 179;
Best Local Similarity 44.8%; Pred. No. 6.1;
Matches 13; Conservative 4; Mismatches 11; Indels 1

QY 4 FWGDTLNCWMLSAFSRYARCLAEHGDGPT 32
|||::: |||: ||| | ||
Db 87 FWCDTIHGVYPMLSRSLRQVAEGKD-PT 114

RESULT 15

Q924K1	PRELIMINARY;	PRT;	328 AA.
ID	Q924K1		
AC	Q924K1;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Aryl-hydrocarbon interacting protein-like 1.		
DE	Aryl-hydrocarbon interacting protein-like 1.		
GN	Alp1.		
GN	Mus musculus (Mouse).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=21313649; PubMed=11420621;		
RX	Sohocki M.M., Sullivan L.S., Tirpak D.L., Daiger S.P.;		
RA	"Comparative analysis of aryl-hydrocarbon receptor interacting		
RT	protein-like 1 (Alp1), a gene associated with inherited retinal		
RT	disease in humans."		
RL	Mamm. Genome 12:566-568(2001).		
RL	EMBL; AF296412; AAk7956.1; "		
RT	MGM; MGI:2148800; Alp1.		
DR	InterPro; IPR001440; TPR.		
DR	Pfam; PF00515; TPR; 2.		
DR	SEQUENCE 328 AA; 38303 MW; 5D34AAE530A325A7 CRC64;		

Query Match 27.98; Score 53.5; DB 11; Length 328;
Best Local Similarity 44.88; Pred. No. 11;
Matches 13; Conservative 4; Mismatches 11; Indels 1;
Gaps 1;

QY 4 FWGDTLNCWMLSAFSRYARCLAEAGHDGPT 32

Thu May 29 16:01:03 2003

us-09-989-481-4.rspt

Page 6

Db 87 FWCDTHTGVIPMLSRSLRQVAEGKD-PT 114

Search completed: May 29, 2003, 15:45:02
Job time : 82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 15:34:05 ; Search time 22 seconds
(without alignments)
62.214 Million cell updates/sec

Title: US-09-989-481-4
Perfect score: 192
Sequence: 1 LGTFWGTLCNWLMSAFSRYARCLAEHGDPTQ 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	30.7	910	1 IMB2_SCHPO	014089 schizosacch
2	53.5	27.9	328	1 APLP_RAT	09199 rattus norv
3	53.5	27.9	384	1 APLP_HUMAN	09299 homo sapien
4	52	27.1	563	1 LIPN_GEOCN	P17573 geotrichum
5	51	26.6	418	1 CGAL_XENLA	P18606 xenopus lae
6	51	26.6	474	1 ASCB_ECOLI	P24240 escherichia
7	51	26.6	1207	1 EGF_HUMAN	P01133 homo sapien
8	49.5	25.8	355	1 GBA2_NEUCR	Q05424 neurospora
9	49.5	25.8	847	1 ENV_HV1S1	P19550 human immun
10	49	25.5	282	1 APAH_BURMA	09aev8 burkholderi
11	49	25.5	282	1 APAH_BURPS	069115 burkholderi
12	48	25.0	229	1 PEPE_ECOLI	P32666 escherichia
13	48	25.0	472	1 HEAD_BPGAL	09fz7 bacterioph
14	48	25.0	1959	1 AGRI_RAT	P25304 rattus norv
15	47	24.5	690	1 PPK_PSEAE	09s646 pseudomonas
16	47	24.5	3433	1 POLG_KUNJM	P14335 k genome po
17	46.5	24.2	729	1 NARE_SYNP7	P39458 synecococc
18	46.5	24.2	895	1 ODPI_ALCEU	059097 alcaligenes
19	46	24.0	182	1 C560_CAEEL	P41956 caenorhabdi
20	46	24.0	184	1 C560_CAEER	P41956 caenorhabdi
21	46	24.0	361	1 COOH_RHORI	P31895 rhodospiril
22	46	24.0	449	1 HEAD_BPGO3	03088 bacterioph
23	46	24.0	1550	1 GLTB_SYNY3	P55037 synecocyst
24	46	24.0	1822	1 ITB4_HUMAN	P16144 homo sapien
25	45.5	23.7	126	1 YF81_XYLFA	09p82 xyella fas
26	45.5	23.7	614	1 VAAL_DROME	P48602 drosophila
27	45	23.4	114	1 RSN_MOUSE	099p87 mus musculu
28	45	23.4	158	1 NEU4_ONCKE	P16042 oncorhynch
29	45	23.4	334	1 GBLP_ORYSA	P49027 oryza sativ
30	45	23.4	404	1 VE2_HPV60	080944 human papil
31	45	23.4	698	1 PPK_XYLFA	09pac7 xyella fas
32	45	23.4	1597	1 SOL_DROME	P27398 drosophila
33	45	23.4	3038	1 TRIO_HUMAN	075962 homo sapien

RESULT 1
IMB2_SCHPO
ID IMB2_SCHPO STANDARD: PRT: 910 AA.
AC 014089;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative importin beta-2 subunit (Karyopherin beta-2 subunit)
DE (Importin 104) (Transportin) (TRN).
GN SPAC2F3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA WOOD V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gatties S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson A.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tilvey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vansteenkiste E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -1- FUNCTION: REQUIRED FOR IMPORT OF MRNA BINDING PROTEINS. BINDS TO
CC -1- NUCLEOPORINS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE IMPORTIN BETA FAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMPORTIN N-TERMINAL DOMAIN.
CC -1- SIMILARITY: CONTAINS 9 HEAT REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

P06935 w genome po
Q29487 felis silve
Q08869 pasteurella
P80147 sus scrofa
P50534 rattus norv
P31404 bos taurus
P38606 homo sapien
P50516 mus musculu
Q29048 sus scrofa
P15145 sus scrofa
P51527 meleagris g
P00746 homo sapien

ALIGNMENTS

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z99165; CAB16272.1; -

DR HSSP: Q92973; LOBK.

DR PROSITE: PS0077; HEAT_REPEAT; FALSE_NEG.

DR PROSITE: PS0166; IMPORTIN_B_NT; FALSE_NEG.

KW Hypothetical protein; Transport; Protein transport; Repeat.

FT DOMAIN 34 122 IMPORTIN N-TERMINAL.

FT REPEAT 127 164 HEAT 1.

FT REPEAT 174 211 HEAT 2.

FT REPEAT 299 336 HEAT 3.

FT REPEAT 410 447 HEAT 4.

FT REPEAT 451 488 HEAT 5.

FT REPEAT 497 534 HEAT 6.

FT REPEAT 538 575 HEAT 7.

FT REPEAT 569 608 HEAT 8.

FT REPEAT 639 679 HEAT 9.

FT DOMAIN 366 385 ASP/GLU-RICH (ACIDIC).

FT SEQUENCE 910 AA; 101718 MW; 4939CD9B09877208 CRC64;

Query Match 30.7%; Score 59; DB 1; Length 910;

Best Local Similarity 40.9%; Pred. No. 2;

Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 8 TLNCWMLSAFSAFYARCLAEHGD 29

DB 473 TITCWTGLGRYSKWASCLESEED 494

RESULT 2

AIPL_RAT

ID AIPL_RAT STANDARD; PRT; 328 AA.

AC Q9JLG9;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Aryl-hydrocarbon interacting protein-like 1.

GN AIPL1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-20082814; PubMed-10615133;

RA Sothocki M.M., Bowne S.J., Sullivan L.S., Blackshaw S., Cepko C.L.,

RA Payne A.M., Bhattacharya S.S., Khalil Q., Birch D.G.,

RA Harrison W.R., Elder F.F.B., Heckenlively J.R., Daiger S.P.,

RA "Mutations in a novel photoreceptor-pineal gene on 17p cause Leber

RA congenital amaurosis."

RL Nat. Genet. 24:79-83(2000).

CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN RETINA.

CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.

CC -1- SIMILARITY: CONTAINS 2 TPR REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AF180340; AAF26707.1; -

DR InterPro: IPR001179; FKBP_PPIase.

DR InterPro: IPR001440; TPR.

DR Pfam: PF00515; TPR; 2.

DR PROSITE: PS00453; FKBP_PPIASE_1; FALSE_NEG.

DR PROSITE: PS00454; FKBP_PPIASE_2; FALSE_NEG.

DR PROSITE: PS00509; FKBP_PPIASE_3; FALSE_NEG.

DR Repeat; TPR repeat.

KW

FT DOMAIN 53 145 PPIASE, FKBP-TYPE.

FT REPEAT 230 263 TPR 1.

FT REPEAT 264 297 TPR 2.

SQ SEQUENCE 328 AA; 38294 MW; E9BC3A4084F64A0E CRC64;

Query Match 27.9%; Score 53.5; DB 1; Length 328;

Best Local Similarity 44.8%; Pred. No. 4.1;

Matches 13; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

QY 4 FWGDTLNCWMLSAFSAFYARCLAEHGDGPT 32

DB 87 FWCDFTHGTGVPMLSRSLRQVAEGKD-PT 114

RESULT 3

AIPL_HUMAN

ID AIPL_HUMAN STANDARD; PRT; 384 AA.

AC Q9NZN9;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Aryl-hydrocarbon interacting protein-like 1.

GN AIPL1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., VARIANT HIS-90, AND VARIANT LCA4 ARG-239.

RX MEDLINE-20082814; PubMed-10615133;

RA Sothocki M.M., Bowne S.J., Sullivan L.S., Blackshaw S., Cepko C.L.,

RA Payne A.M., Bhattacharya S.S., Khalil Q., Birch D.G.,

RA Harrison W.R., Elder F.F.B., Heckenlively J.R., Daiger S.P.,

RA "Mutations in a novel photoreceptor-pineal gene on 17p cause Leber

RA congenital amaurosis."

RL Nat. Genet. 24:79-83(2000).

CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN RETINA.

CC -1- DISEASE: DEFECTS IN AIPL1 ARE A CAUSE OF LEBER CONGENITAL

CC AMAUROSIS TYPE 4 (LCA4), A DISEASE CHARACTERIZED BY TOTAL

CC BLINDNESS OR GREATLY IMPAIRED VISION WITH LOSS OF CENTRAL VISION.

CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.

CC -1- SIMILARITY: CONTAINS 2 TPR REPEATS.

CC -1- DATABASE: NAME-Mutations of the AIPL1 gene.

CC NOTE-Retina International's Scientific Newsletter;

CC WWW="http://www.retina-international.com/sci-news/aiplmut.htm".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AF180472; AAF26708.1; -

DR Genew; HGNC:359; AIPL1.

DR MIM; 604392; -

DR InterPro: IPR001179; FKBP_PPIase.

DR InterPro: IPR001440; TPR.

DR Pfam: PF00515; TPR; 2.

DR PROSITE: PS00453; FKBP_PPIASE_1; FALSE_NEG.

DR PROSITE: PS00454; FKBP_PPIASE_2; FALSE_NEG.

DR PROSITE: PS00509; FKBP_PPIASE_3; FALSE_NEG.

DR Repeat; TPR repeat; Disease mutation; Vision.

FT DOMAIN 53 145 PPIASE, FKBP-TYPE.

FT REPEAT 230 263 TPR 1.

FT REPEAT 264 297 TPR 2.

FT VARIANT 90 90 D -> H.

FT VARIANT 239 239 /FTIG-VAR_010140.

FT C -> R (IN LCA4).

FT /FTIG-VAR_010139.

FT SEQUENCE 384 AA; 43596 MW; B116AA0724BAF9EA CRC64;

Query Match 27.9%; Score 53.5; DB 1; Length 384;
 Best Local Similarity 40.0%; Pred. No. 4.7;
 Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Oy 4 FWDGTLNCWMLSAFSAFYARCLAGHDGPTQ 33
 |||::: : |||::: |||::: |||:::
 Db 87 FWCDTIHTGVYPILSRSLRQMAQCKD-PTE 115

RESULT 4

LIPI_GEOCN STANDARD; PRT; 563 AA.
 AC P17573;
 DC 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Lipase 1 precursor (EC 3.1.1.3).
 OS Geotrichum candidum (Oospora lactis).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Dipodascaceae; Galactomyces.
 OX NCBI_Taxid-273117;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-ATCC 34614;
 RX MEDLINE-90110016; PubMed-2481674;
 RA Shimada Y., Sugihara A., Tomimaga Y., Iizumi T., Tsunasawa S.;
 RT "cDNA molecular cloning of Geotrichum candidum lipase.";
 RL J. Biochem. 106:383-388(1989).
 RN [2]
 RP SIMILARITY TO CARBOXYLESTERASES.
 RX MEDLINE-90328988; PubMed-2115773;
 RA Slabas A.R., Windust J., Sidebottom C.M.;
 RT "Does sequence similarity of human choline esterase, Torpedo
 acetylcholine esterase and Geotrichum candidum lipase reveal the
 active site serine residue?";
 RL Biochem. J. 269:279-280(1990).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE-91287805; PubMed-2062369;
 RA Schrag J.D., Li Y., Wu S., Cysler M.;
 RT "Ser-His-Glu triad forms the catalytic site of the lipase from
 Geotrichum candidum.";
 RL Nature 351:761-765(1991).
 CC -1- FUNCTION: THE EXTRACELLULAR LIPASE PRODUCED BY G. CANDIDUM
 HYDROLYZES ALL ESTER BONDS IN TRIGLYCERIDE AND DISPLAYS A HIGH
 AFFINITY FOR TRIOLEIN.
 CC -1- CATALYTIC ACTIVITY: triacylglycerol + H(2)O = diacylglycerol + a
 fatty acid anion.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 DR PIR: JQ0022; ACUGC.
 DR PDB: 1THG; 3I-OCT-93.
 DR InterPro: IPR002018; CarbesteraseB.
 DR InterPro: IPR000379; Ser_estrs_site.
 DR Pfam: PF00135; Coesterase; 1.
 DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
 KW Hydrolyase; Lipid degradation; Glycoprotein; Signal; 3D-structure.
 FT SIGNAL 1
 FT CHAIN 20 563
 FT MOD_RES 20 20 LIPASE 1.
 FT ACT_SITE 20 20 PYRROLIDONE CARBOXYLIC ACID.
 FT ACT_SITE 236 236
 FT ACT_SITE 373 373
 FT ACT_SITE 482 482
 FT DISULFID 80 124
 FT DISULFID 295 307
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 23 26
 FT TURN 27 29
 FT STRAND 30 33

FT STRAND 37
 FT TURN 38
 FT STRAND 40
 FT HELIX 47
 FT TURN 54
 FT TURN 56
 FT TURN 57
 FT TURN 69
 FT STRAND 70
 FT STRAND 71
 FT STRAND 72
 FT STRAND 74
 FT STRAND 77
 FT STRAND 79
 FT STRAND 85
 FT HELIX 96
 FT HELIX 98
 FT HELIX 101
 FT HELIX 104
 FT TURN 113
 FT TURN 114
 FT STRAND 120
 FT STRAND 126
 FT STRAND 132
 FT TURN 133
 FT TURN 134
 FT TURN 137
 FT STRAND 141
 FT STRAND 147
 FT TURN 151
 FT TURN 152
 FT HELIX 159
 FT HELIX 163
 FT HELIX 171
 FT TURN 173
 FT STRAND 177
 FT TURN 181
 FT HELIX 186
 FT HELIX 194
 FT TURN 199
 FT TURN 200
 FT TURN 202
 FT HELIX 204
 FT TURN 223
 FT STRAND 224
 FT STRAND 225
 FT TURN 235
 FT TURN 236
 FT HELIX 237
 FT TURN 250
 FT TURN 251
 FT TURN 255
 FT TURN 256
 FT STRAND 259
 FT STRAND 263
 FT STRAND 267
 FT HELIX 285
 FT TURN 294
 FT TURN 298
 FT TURN 299
 FT TURN 301
 FT HELIX 310
 FT HELIX 313
 FT TURN 329
 FT TURN 331
 FT HELIX 334
 FT TURN 337
 FT TURN 351
 FT HELIX 355
 FT TURN 357
 FT STRAND 365
 FT TURN 370
 FT TURN 371
 FT STRAND 372
 FT STRAND 373
 FT TURN 377
 FT TURN 381
 FT TURN 382
 FT TURN 383
 FT HELIX 387
 FT TURN 397
 FT TURN 400
 FT HELIX 403
 FT HELIX 412
 FT HELIX 419
 FT TURN 423
 FT TURN 424
 FT TURN 426
 FT HELIX 435
 FT TURN 447
 FT TURN 447
 FT HELIX 447
 FT TURN 457
 FT TURN 459
 FT STRAND 460
 FT STRAND 463
 FT TURN 470
 FT TURN 474
 FT TURN 476
 FT STRAND 481
 FT TURN 482
 FT TURN 485
 FT HELIX 486
 FT TURN 490
 FT TURN 491
 FT TURN 495
 FT TURN 496
 FT HELIX 497
 FT TURN 513
 FT TURN 514


```

QY 1 LGTFWDTLNCWMLSAFSRYARCCLAECHDG 30
DB 141 LVTEGSRNRKLVFFSRYARTCFEADG 170

RESULT 7
EGF_HUMAN
ID EGF_HUMAN STANDARD; PRT; 1207 AA.
AC P01133;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pro-epidermal growth factor precursor (EGF) [Contains: Epidermal
DE growth factor (Urogastrone)].
GN EGF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA MEDLINE=87066721; PubMed=3491360;
RA Bell G.I., Fong N.M., Stempien M.M., Wormsted M.A., Caput D.,
RA Ku L., Urdea M.S., Rall L.B., Sanchez-Pescador R.;
RA "Human epidermal growth factor precursor: cDNA sequence, expression
RA in vitro and gene organization.";
RA Nucleic Acids Res. 14:8427-8446(1986).
RN [2]
RP SEQUENCE OF 971-1023.
RX MEDLINE=77117897; PubMed=300079;
RX Gregory H., Preston B.M.;
RX "The primary structure of human urogastrone.";
RX Int. J. Pept. Protein Res. 9:107-118(1977).
RN [3]
RP SEQUENCE OF 971-1023.
RX MEDLINE=89391964; PubMed=2789514;
RX Furuya M., Akashi S., Hirayama K.;
RX "The primary structure of human EGF produced by genetic engineering,
RX studied by high-performance tandem mass spectrometry.";
RX Biochem. Biophys. Res. Commun. 163:1100-1106(1989).
RN [4]
RP STRUCTURE BY NMR OF EGF.
RX MEDLINE=92395667; PubMed=1522591;
RX Hommel U., Harvey T.S., Driscoll P.C., Campbell I.D.;
RX "Human epidermal growth factor. High resolution solution structure
RX and comparison with human transforming growth factor alpha.";
RX J. Mol. Biol. 227:271-282(1992).
CC -1- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
CC FIBROBLASTS IN CELL CULTURE.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X04571; CAA28240.1; -.
CC PIR; A25531; EGHU
CC HSP; P01132; EGF.
CC Genew; HGNC:3229; EGF.
CC MIM; 131530; -.
CC InterPro; IPR000152; Asx_hydroxyl.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001336; EGF_1.
CC InterPro; IPR01881; EGF_Ca.
CC InterPro; IPR000033; Ldl_receptor_rep.

```

```

DR Pfam; PF00008; EGF; 9.
DR Pfam; PF00058; ldl_recept_b; 7.
DR PRINTS; PRO0009; EGF_TGF.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00001; EGF_Like; 7.
DR SMART; SM00135; LY; 8.
DR PROSITE; PS00010; ASX_HYDROXYL; 3.
DR PROSITE; PS00023; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 7.
DR PROSITE; PS01187; EGF_CA; 3.
DR EGF-like domain; Repeat; Growth factor; Transmembrane; Glycoprotein;
KW Signal; Polymorphism.
FT SIGNAL 1 22
FT CHAIN 23 1207
FT CHAIN 971 1023
FT DOMAIN 23 1032
FT TRANSMEM 1033 1053
FT DOMAIN 1054 1207
FT DOMAIN 314 355
FT DOMAIN 356 396
FT DOMAIN 397 437
FT DOMAIN 435 477
FT DOMAIN 741 781
FT DOMAIN 831 869
FT DOMAIN 870 911
FT DOMAIN 912 952
FT DOMAIN 972 1013
FT DISULFID 318 330
FT DISULFID 325 339
FT DISULFID 341 354
FT DISULFID 360 371
FT DISULFID 367 380
FT DISULFID 382 395
FT DISULFID 401 412
FT DISULFID 408 421
FT DISULFID 423 436
FT DISULFID 439 451
FT DISULFID 447 461
FT DISULFID 463 476
FT DISULFID 745 756
FT DISULFID 752 765
FT DISULFID 767 780
FT DISULFID 835 846
FT DISULFID 840 855
FT DISULFID 857 868
FT DISULFID 874 888
FT DISULFID 881 897
FT DISULFID 899 910
FT DISULFID 916 929
FT DISULFID 923 938
FT DISULFID 940 951
FT DISULFID 976 990
FT DISULFID 984 1001
FT DISULFID 1003 1012
FT CARBOHYD 38 38
FT CARBOHYD 104 104
FT CARBOHYD 117 117
FT CARBOHYD 148 148
FT CARBOHYD 324 324
FT CARBOHYD 404 404
FT CARBOHYD 596 596
FT CARBOHYD 815 815
FT CARBOHYD 926 926
FT VARIANT 708 708
FT SQ SEQUENCE 1207 AA; 133945 MW; D627DC828EF782E9 CRC64;
Query Match 26.6%; Score 51; DB 1; Length 1207;
Best Local Similarity 56.2%; Pred. No. 32;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 18 SRYARCCLAECHDGPTQ 33
| | | | | | | | | |

```

```

Db      841 SMYARCISEGDATECQ 856

RESULT 8
GBA2_NEUCR
ID      GBA2_NEUCR      STANDARD;      PRT;      355 AA.
AC      Q05424; Q9URK0;
DT      01-OCT-1994 (Rel. 30, Created)
DT      01-OCT-1994 (Rel. 30, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Guanine nucleotide-binding protein alpha-2 subunit (GP2-alpha).
GN      GNA-2.
OS      Neurospora crassa.
OC      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC      Sordariales; Sordariaceae; Neurospora.
OX      NCBI_TaxID=5141;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=74-OR23-1A;
RX      MEDLINE=93315452; PubMed=8325859;
RA      Borkovich K.A., Turner G.E.;
RT      "Identification of a G protein alpha subunit from Neurospora crassa
RT      that is a member of the Gi family.";
RL      J. Biol. Chem. 268:14805-14811(1993).
[2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=74-OR23-1A;
RX      MEDLINE=97432794; PubMed=9286674;
RA      Baasiri R.A., Lu X., Rowley P.S., Turner G.E., Borkovich K.A.;
RT      "Overlapping functions for two G protein alpha subunits in Neurospora
RT      crassa.";
RL      Genetics 147:137-145(1997).
CC      -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC      INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE
CC      SIGNALING SYSTEMS.
CC      -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC      THE ALPHA CHAIN CONTAINS THE GUANINE NUCLEOTIDE BINDING SITE.
CC      -1- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY, SUBFAMILY 3 (G(Q)).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L11452; AAA02559.1;
DR      EMBL; AF004846; AAD01207.1;
DR      InterPro; IPR001019; Gprotein_alpha.
DR      Pfam; PF00503; G-alpha; 1.
DR      PRINTS; PR00318; GPROTEINA.
DR      ProDom; PD000281; Gprotein_alpha; 1.
DR      SMART; SM00275; G-alpha; 1.
KW      GTP-binding; Transducer; Multigene family.
FT      NP_BIND 41 48 GTP (BY SIMILARITY).
FT      NP_BIND 201 205 GTP (BY SIMILARITY).
FT      NP_BIND 270 273 GTP (BY SIMILARITY).
FT      CONFLICT 19 20 EL -> DV (IN REF. 2).
SQ      SEQUENCE 355 AA; 41361 MW; 61733B89EABB7409 CRC64;

Query Match      25.8%; Score 49.5; DB 1; Length 355;
Best Local Similarity 44.0%; Pred. No. 15;
Matches 11; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

QY      7 DTLCNWM-LSAFSRVYARCLAEHGHDG 30
      : : : : : : : : : : : : : : : :
Db      217 ENVNCLFLVAISGYDQCLVEDKDG 241

RESULT 9
ENV_HV1S1
ID      ENV_HV1S1      STANDARD;      PRT;      847 AA.

```

```

AC      P19550;
DT      01-FEB-1991 (Rel. 17, Created)
DT      01-FEB-1991 (Rel. 17, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Envelope glycoprotein GP160 precursor (Contains: Exterior membrane
DE      glycoprotein (GP120); Transmembrane glycoprotein (GP41)).
GN      ENV.
OS      Human immunodeficiency virus type 1 (SF162 isolate) (HIV-1).
OC      Viruses; Retrovird viruses; Retroviridae; Lentivirus.
OX      NCBI_TaxID=11691;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90347835; PubMed=2384920;
RA      Cheng-Mayer C., Quiroga M., Tung J.W., Dina D., Levy J.;
RT      "Viral determinants of human immunodeficiency virus type 1 T-cell or
RT      macrophage tropism, cytopathogenicity, and CD4 antigen modulation.";
RL      J. Virol. 64:4390-4398(1990).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M65024; AAA45072.1;
DR      HIV; M38428; ENVSSF162.
DR      InterPro; IPR000328; Env_GP41.
DR      Pfam; PF00516; GP120; 1.
DR      Pfam; PF00517; GP41; 1.
KW      AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
KW      Signal.
FT      SIGNAL 1 29
FT      CHAIN 30 502
FT      CHAIN 503 847
FT      DISULFID 53 73
FT      DISULFID 118 203
FT      DISULFID 125 194
FT      DISULFID 130 155
FT      DISULFID 216 245
FT      DISULFID 226 237
FT      DISULFID 234 328
FT      DISULFID 374 435
FT      DISULFID 381 408
FT      CARBOHYD 87 87
FT      CARBOHYD 135 135
FT      CARBOHYD 134 134
FT      CARBOHYD 186 186
FT      CARBOHYD 195 195
FT      CARBOHYD 232 232
FT      CARBOHYD 239 239
FT      CARBOHYD 260 260
FT      CARBOHYD 274 274
FT      CARBOHYD 293 293
FT      CARBOHYD 299 299
FT      CARBOHYD 329 329
FT      CARBOHYD 336 336
FT      CARBOHYD 332 352
FT      CARBOHYD 382 382
FT      CARBOHYD 388 388
FT      CARBOHYD 392 392
FT      CARBOHYD 398 398
FT      CARBOHYD 401 401
FT      CARBOHYD 438 438
FT      CARBOHYD 454 454
FT      CARBOHYD 602 602
FT      CARBOHYD 607 607
FT      CARBOHYD 616 616
FT      CARBOHYD 628 628
SQ      SEQUENCE 847 AA; 96135 MW; 0A901317FD7FF2AB CRC64;

```


RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
 RL Nature 409:529-533(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 CC -I- FUNCTION: Hydrolyzes dipeptides containing N-terminal aspartate
 residues. May play a role in allowing the cell to use peptide
 aspartate to spare carbon otherwise required for the synthesis of
 the aspartate family of amino acids.
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S51.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U00006; AAC43115.1; -
 DR EMBL: AE000475; AAC76991.1; -
 DR EMBL: AE005634; AG59213.1; -
 DR EMBL: AP002567; BAB38362.1; -
 DR HSSP: P36936; 1FYE.
 DR MEROPS: S51.001; -
 DR EcoGene: EG11920; pepE.
 DR InterPro: IPR005320; Peptidase_S51.
 DR Pfam: PF03575; Peptidase_S51; 1.
 KW Hydrolyase; Serine protease; Dipeptidase; Complete proteome.
 FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 135 135 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 157 157 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 229 AA; 24570 MW; 53D4D8395DFC63FD CRC64;
 Query Match 25.0%; Score 48; DB 1; Length 229;
 Best Local Similarity 37.0%; Pred. No. 16;
 Matches 10; Conservative 3; Mismatches 14; Indels 0; Gaps 0;
 QY 7 DTLCNWMLSAFSRVYARCLAEHGDPQTQ 33
 Db 145 DALNPLQLINPHFTNALPEGHGETR 171
 RESULT 13
 HEAD_BPGAL STANDARD; PRT; 472 AA.
 AC Q9FZW7;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Major head protein.
 GN 8.
 OS Bacteriophage GA-1.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
 OC phi-29-like viruses.
 OX NCBI_TaxID=12345;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Meijer W.J.J., Horcajadas J.A., Salas M.;
 RT "The phi29 family of phages";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X96987; CAC21529.1; -
 SQ SEQUENCE 472 AA; 53022 MW; 3104821153B1C4C2 CRC64;
 Query Match 25.0%; Score 48; DB 1; Length 472;
 Best Local Similarity 33.3%; Pred. No. 32;
 Matches 8; Conservative 6; Mismatches 10; Indels 0; Gaps 0;
 QY 2 GTFWGDTLCNWMLSAFSRVYARCLAE 25
 Db 318 GWYNYLHWQVLSISREANAVA 341
 RESULT 14
 AGRI_RAT STANDARD; PRT; 1959 AA.
 ID AGRI_RAT
 AC P25304; Q63034;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Agrin precursor.
 GN AGRN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE OF 1-1779 AND 1799-1959 FROM N.A.
 RC TISSUE=Embryonic spinal cord;
 RX MEDLINE=91222570; PubMed=1851019;
 RA Rupp F., Payan D.G., Magill-Solc C., Cowan D.M., Scheller R.H.;
 RT "Structure and expression of a rat agrin";
 RL Neuron 6:811-823(1991).
 RN [2]
 RP SEQUENCE OF 1777-1801 FROM N.A.
 RX MEDLINE=92407628; PubMed=1326608;
 RA Rupp F., Oezcelik T., Linial M., Peterson K., Francke U., Scheller R.;
 RT "Structure and chromosomal localization of the mammalian agrin gene";
 RL J. Neurosci. 12:3535-3544(1992).
 CC -I- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
 AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
 ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.
 CC -I- SUBUNIT: BINDS TO LAMININ.
 CC -I- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
 JUNCTION.
 CC -I- ALTERNATIVE PRODUCTS: AT LEAST 5 ISOFORMS; 1 (SHOWN HERE), 2, 3, 4
 AND 5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
 ACETYLCHOLINE RECEPTOR CLUSTERING ACTIVITY.
 CC -I- TISSUE SPECIFICITY: EMBRYONIC NERVOUS SYSTEM AND MUSCLE.
 CC -I- DEVELOPMENTAL STAGE: MORE ABUNDANT EARLY IN DEVELOPMENT.
 CC -I- PTM: CONTAINS HEPARAN SULFATE CHAINS AS WELL AS N-LINKED AND O-
 LINKED OLIGOSACCHARIDES (BY SIMILARITY).
 CC -I- SIMILARITY: CONTAINS 9 KAZAL-LIKE DOMAINS.
 CC -I- SIMILARITY: CONTAINS 2 LAMININ EGF-LIKE DOMAINS.
 CC -I- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
 CC -I- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -I- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
 CC -I- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-18 OR MET-24 IS THE
 INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 15:41:05 ; Search time 40 seconds
(without alignments)
79.311 Million cell updates/sec

Title: US-09-989-481-4
Perfect score: 192
Sequence: 1 LGTFWDTLNCWMLSAFSAFYARCLAEHGDPQ 33
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	30.7	910	2 T38539	probable Importin
2	56.5	29.4	223	2 D70760	hypothetical prote
3	55	28.6	694	2 T52574	cyclic nucleotide
4	54	28.1	287	2 T3197	yeast atp12 protei
5	54	28.1	316	2 C87318	hypothetical prote
6	52	27.1	255	2 G87668	conserved hypothet
7	52	27.1	533	2 T01864	hypothetical prote
8	52	27.1	544	2 S41090	triacylglycerol li
9	52	27.1	563	1 ACGUGC	triacylglycerol li
10	51.5	26.8	1150	2 S58775	mypl protein - smu
11	51	26.6	90	2 S24248	Ig heavy chain v r
12	51	26.6	101	2 S24257	Ig heavy chain v r
13	51	26.6	105	2 S24249	Ig heavy chain v r
14	51	26.6	109	2 S24254	Ig heavy chain v r
15	51	26.6	109	2 S24253	Ig heavy chain v r
16	51	26.6	110	2 S24250	Ig heavy chain v r
17	51	26.6	113	2 S24247	Ig heavy chain v r
18	51	26.6	373	2 G85355	nodulin-like prote
19	51	26.6	418	2 S11678	cyclin A - African
20	51	26.6	472	2 D75464	probable zinc meta
21	51	26.6	474	2 D91075	6-phospho-beta-glu
22	51	26.6	474	2 C85920	6-phospho-beta-glu
23	51	26.6	474	2 H65051	6-phospho-beta-glu
24	51	26.6	1207	1 EGHU	epidermal growth f
25	50.5	26.3	537	2 T05816	hypothetical prote
26	50	26.0	726	2 T51519	cyclic nucleotide
27	49.5	25.8	355	2 T50479	G protein alpha ch
28	49.5	25.8	408	2 T38386	hypothetical wd-40
29	49	25.5	117	2 T46430	hypothetical prote

30 49 25.5 414 2 H86736 rod-shape determin
31 49 25.5 709 2 D82825 conserved hypothet
32 48.5 25.3 293 2 D95919 hypothetical prote
33 48.5 25.3 506 2 T35261 probable metallopro
34 48 25.0 135 2 H71051 hypothetical prote
35 48 25.0 229 2 D65209 peptidase E - Esch
36 48 25.0 229 2 C91246 peptidase E (impor
37 48 25.0 229 2 A86094 peptidase E (impor
38 48 25.0 353 2 AC3320 integral membrane
39 48 25.0 1959 1 AGRT agrin - rat
40 47.5 24.7 339 2 T05217 protein sam2B - sp
41 47.5 24.7 467 2 G83592 adenosylmethionine
42 47.5 24.7 477 2 T05202 pectinesterase hom
43 47 24.5 102 2 S24260 Ig heavy chain v r
44 47 24.5 115 2 T31781 hypothetical prote
45 47 24.5 241 2 C82852 competence protein

ALIGNMENTS

RESULT 1

T38539

Probable importin beta-2 subunit (transportin) - fission yeast (Schizosaccharomyces p.

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T38539

R:Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, September 1997

A:Reference number: Z31748

A:Accession: T38539

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-910 <OLI>

A:Cross-references: EMBL:Z99165; PIDN:CAB16272.1; GSPDB:GN00066; SPDB:SPAC2F3.06c

A:Experimental source: strain 972h-; cosmid c2f3

C:Genetics:

A:Gene: SPDB:SPAC2F3.06c

A:Map position: 1

A:Introns: 36/3

Query Match 30.7%; Score 59; DB 2; Length 910;
Best Local Similarity 40.9%; Pred. No. 3.7;
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 8 TLNCWMLSAFSAFYARCLAEHGD 29

DB 473 TTTCWTLKRYKWSCLSEED 494

RESULT 2

D70760

hypothetical protein RV2014 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: D70760

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, K.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: D70760

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-223 <COL>

A:Cross-references: GE:Z74025; GB:AL123456; NID:g3261586; PIDN:CAA98415.1; PID:el12999

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV2014

Query Match 29.4%; Score 56.5; DB 2; Length 223;

RESULT 12

S24257
Ig heavy chain V region (VH26-DXP1-JH4) - human
C:Species: Homo sapiens (man)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
C:Accession: S24257
R:Stewart, A.K.; Huang, C.; Stollar, B.D.; Schwartz, R.S.
submitted to the EMBL Data Library, June 1992
A:Description: A single VH gene predominates in the rearranged and expressed human B cell
A:Reference number: S24247
A:Accession: S24257
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-101 <STE>
A:Cross-references: EMBL:X67065; PIDN:CAA47450.1; PID:g38388
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:11-93/Domain: immunoglobulin homology <IMM>

Query Match 26.6%; Score 51; DB 2; Length 101;
Best Local Similarity 52.2%; Pred. No. 5.2;
Matches 12; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 LGTFWG--DTLNCWMLSAFSRYA 21
||||| :||| : ||||
Db 6 LGTAWGVPETLLCSLWFTFSYA 28

RESULT 13

S24249
Ig heavy chain V region (VH26-DN1-DXP1-JH4) - human
C:Species: Homo sapiens (man)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 30-May-1997
C:Accession: S24249
R:Stewart, A.K.; Huang, C.; Stollar, B.D.; Schwartz, R.S.
submitted to the EMBL Data Library, June 1992
A:Description: A single VH gene predominates in the rearranged and expressed human B cell
A:Reference number: S24247
A:Accession: S24249
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-105 <STE>
A:Cross-references: EMBL:X67070
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:10-92/Domain: immunoglobulin homology <IMM>

Query Match 26.6%; Score 51; DB 2; Length 105;
Best Local Similarity 52.2%; Pred. No. 5.5;
Matches 12; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 LGTFWG--DTLNCWMLSAFSRYA 21
||||| :||| : ||||
Db 5 LGTAWGVPETLLCSLWFTFSYA 27

RESULT 14

S24254
Ig heavy chain V region (VH26-DXP2-JH4) - human
C:Species: Homo sapiens (man)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
C:Accession: S24254
R:Stewart, A.K.; Huang, C.; Stollar, B.D.; Schwartz, R.S.
submitted to the EMBL Data Library, June 1992
A:Description: A single VH gene predominates in the rearranged and expressed human B cell
A:Reference number: S24247
A:Accession: S24254
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-109 <STE>
A:Cross-references: EMBL:X67062

C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:15-97/Domain: immunoglobulin homology <IMM>

Query Match 26.6%; Score 51; DB 2; Length 109;
Best Local Similarity 52.2%; Pred. No. 5.7;
Matches 12; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 LGTFWG--DTLNCWMLSAFSRYA 21
||||| :||| : ||||
Db 10 LGTAWGVPETLLCSLWFTFSYA 32

RESULT 15

S24253
Ig heavy chain V region (VH26-DLR4-JH6) - human
C:Species: Homo sapiens (man)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
C:Accession: S24253
R:Stewart, A.K.; Huang, C.; Stollar, B.D.; Schwartz, R.S.
submitted to the EMBL Data Library, June 1992
A:Description: A single VH gene predominates in the rearranged and expressed human B
A:Reference number: S24247
A:Accession: S24253
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-109 <STE>
A:Cross-references: EMBL:X67061
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: heterotetramer; immunoglobulin
F:12-94/Domain: immunoglobulin homology <IMM>

Query Match 26.6%; Score 51; DB 2; Length 109;
Best Local Similarity 52.2%; Pred. No. 5.7;
Matches 12; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 LGTFWG--DTLNCWMLSAFSRYA 21
||||| :||| : ||||
Db 7 LGTAWGVPETLLCSLWFTFSYA 29

Search completed: May 29, 2003, 15:45:53
Job time : 46 secs

; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76
; LENGTH: 328
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(328)
; OTHER INFORMATION: Mouse AIPL1 Protein
US-09-765-061B-76

Query Match 27.9%; Score 53.5; DB 9; Length 328;
Best Local Similarity 44.8%; Pred. No. 9.1;
Matches 13; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

QY 4 FWGDTLNCWMLSAFSAFYARCLAEHGDGPT 32
|||::: : |||::: |||
DB 87 FWCDDTHGTGVPILSRSLRQMAQKGD-PT 114

RESULT 3

US-09-765-061B-74
; Sequence 74, Application US/09765061B
; Publication No. US20030022165A1
; GENERAL INFORMATION:

; APPLICANT: Board of Regents of the University of Texas System
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal gene 17
; FILE REFERENCE: 96606/16UTL
; CURRENT APPLICATION NUMBER: US/09/765,061B
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Papio anubis
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(372)
; OTHER INFORMATION: Baboon AIPL1 Protein
US-09-765-061B-74

Query Match 27.9%; Score 53.5; DB 9; Length 372;
Best Local Similarity 40.0%; Pred. No. 10;
Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 4 FWGDTLNCWMLSAFSAFYARCLAEHGDGPT 33
|||::: : |||::: |||
DB 87 FWCDDTHGTGVPILSRSLRQMAQKGD-PT 115

RESULT 4

US-09-765-061B-78
; Sequence 78, Application US/09765061B
; Publication No. US20030022165A1
; GENERAL INFORMATION:

; APPLICANT: Board of Regents of the University of Texas System
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal gene 17
; FILE REFERENCE: 96606/16UTL
; CURRENT APPLICATION NUMBER: US/09/765,061B
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 78
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Saimiri sciureus
; FEATURE:
; NAME/KEY: PEPTIDE

; LOCATION: (1)..(372)
; OTHER INFORMATION: Squirrel Monkey AIPL1 Protein
US-09-765-061B-78

Query Match 27.9%; Score 53.5; DB 9; Length 372;
Best Local Similarity 40.0%; Pred. No. 10;
Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 4 FWGDTLNCWMLSAFSAFYARCLAEHGDGPT 33
|||::: : |||::: |||
DB 87 FWCDDTHGTGVPILSRSLRQMAQKGD-PT 115

RESULT 5

US-09-765-061B-72
; Sequence 72, Application US/09765061B
; Publication No. US20030022165A1
; GENERAL INFORMATION:

; APPLICANT: Board of Regents of the University of Texas System
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal gene
; FILE REFERENCE: 96606/16UTL
; CURRENT APPLICATION NUMBER: US/09/765,061B
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(384)
; OTHER INFORMATION: Human AIPL1 Protein
; NAME/KEY: misc_feature
; LOCATION: (322)..(322)
; OTHER INFORMATION: Xaa represents any of the twenty amino acids
US-09-765-061B-72

Query Match 27.9%; Score 53.5; DB 9; Length 384;
Best Local Similarity 40.0%; Pred. No. 11;
Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

QY 4 FWGDTLNCWMLSAFSAFYARCLAEHGDGPT 33
|||::: : |||::: |||
DB 87 FWCDDTHGTGVPILSRSLRQMAQKGD-PT 115

RESULT 6

US-09-765-061B-73
; Sequence 73, Application US/09765061B
; Publication No. US20030022165A1
; GENERAL INFORMATION:

; APPLICANT: Board of Regents of the University of Texas System
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal gene
; FILE REFERENCE: 96606/16UTL
; CURRENT APPLICATION NUMBER: US/09/765,061B
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Pan troglodytes
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(384)
; OTHER INFORMATION: Chimpanzee AIPL1 Protein
US-09-765-061B-73

Query Match 27.9%; Score 53.5; DB 9; Length 384;
Best Local Similarity 40.0%; Pred. No. 11;

	Matches	12;	Conservative	6;	Mismatches	11;	Indels	1;	Gaps	1;
Qy	4	FWGDTLNCWNL	SAFSRYARCLAE	GHGPTQ	33					
		:	:		: : :					
Db	87	FWCDDTHGTGY	PILSRSLRMAAG	KD-PTE	115					

```

RESULT 7
US-09-765-061B-77
; Sequence 77, Application US/09765061B
; Publication No. US20030022165A1
; GENERAL INFORMATION:
; APPLICANT: Board of Regents of the University of Texas System
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal gene 17
; TITLE OF INVENTION: Mutations in a No. US20030022165A1el Photoreceptor-pineal gene 17
; FILE REFERENCE: 96606/16UTL
; CURRENT APPLICATION NUMBER: US/09/765,061B
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Macaca mulatta
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(392)
; OTHER INFORMATION: Rhesus Monkey AIPL1 Protein
US-09-765-061B-77

```

```

Query Match      27.9%; Score 53.5; DB 9; Length 392;
Best Local Similarity 40.0%; Pred. No. 11;
Matches 1; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLCNWMLSAFSRVACRLAEGHDGPTQ 33
   |||||: ||| :||| ||| |||
Db 87 FWCDDTHITGVYPIILSRSLRMAQKGD-PTE 115

```

```

RESULT 8
US-09-764-891-3081
; Sequence 3081, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3081
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (12)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3081

```

```

Query Match      26.68; Score 51; DB 9; Length 56;
Best Local Similarity 57.99; Pred. No. 3;
Matches 11; Conservative 1; Mismatches 5; Indels 1;
Gaps 1;

Qy      9  LNCWMLSAFSRYAR--CLA 25
          ||||| | | | |
Db      25  LNCWHLSCFNHRLRLSCLA 43

```

RESULT 9
US-09-908-711-83

```

: Sequence 83, Application US/09908711
: Patent No. US20020045230A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PA128
: CURRENT APPLICATION NUMBER: US/09/908,711
: CURRENT FILING DATE: 2001-07-20
: PRIOR APPLICATION NUMBER: US01/01360
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,867
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01344
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,892
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01345
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,888
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01329
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,905
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01354
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,891
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01339
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,869
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01340
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,874
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01334
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,898
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01320
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,902
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01329
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,870
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01348
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,882
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01347
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,896
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01348
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,882
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01347
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,864
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01341
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,856
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01336
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: 09/764,868
: PRIOR FILING DATE: 2001-01-17
: PRIOR APPLICATION NUMBER: US01/01312

```

;; PRIOR FILING DATE: 2001-01-17
;; PRIOR APPLICATION NUMBER: 60/179,065
;; PRIOR FILING DATE: 2000-01-31
;; PRIOR APPLICATION NUMBER: 60/180,628
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: 60/209,467
;; PRIOR FILING DATE: 2000-06-07
;; NUMBER OF SEQ ID NOS: 167
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 83
;; LENGTH: 56
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (12)
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-908-711-83

Query Match 26.6%; Score 51; DB 10; Length 56;
Best Local Similarity 57.9%; Pred. No. 3;
Matches 11; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

9 LNCWMLSAFSRYAR--CLA 25
||||| | | | |
25 LNCWHLSCFNHALRLSCLA 43

RESULT 10
US-09-815-242-10270
; Sequence 10270, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

;; APPLICANT: Haselbeck, Robert
;; APPLICANT: Ohlsen, Karl L.
;; APPLICANT: Zyskind, Judith W.
;; APPLICANT: Wall, Daniel
;; APPLICANT: Trawick, John D.
;; APPLICANT: Carr, Grant J.
;; APPLICANT: Yamamoto, Robert T.
;; APPLICANT: Xu, H. Howard

;; TITLE OF INVENTION: Identification of Essential Genes in
;; TITLE OF INVENTION: Prokaryotes

;; FILE REFERENCE: ELITRA.011a
;; CURRENT APPLICATION NUMBER: US/09/815,242

;; CURRENT FILING DATE: 2001-03-21

;; PRIOR APPLICATION NUMBER: 60/191,078

;; PRIOR FILING DATE: 2000-03-21

;; PRIOR APPLICATION NUMBER: 60/206,848

;; PRIOR FILING DATE: 2000-03-23

;; PRIOR APPLICATION NUMBER: 60/207,727

;; PRIOR FILING DATE: 2000-05-26

;; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23

;; PRIOR APPLICATION NUMBER: 60/253,625

;; PRIOR FILING DATE: 2000-11-27

;; PRIOR APPLICATION NUMBER: 60/257,931

;; PRIOR FILING DATE: 2000-12-22

;; PRIOR APPLICATION NUMBER: 60/269,308

;; PRIOR FILING DATE: 2001-02-16

;; NUMBER OF SEQ ID NOS: 14110

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 10270

;; LENGTH: 474

;; TYPE: PRT

;; ORGANISM: Escherichia coli

US-09-815-242-10270

Query Match 26.6%; Score 51; DB 10; Length 474;
Best Local Similarity 43.3%; Pred. No. 31;
Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

1 LGTFWGDTLNCWMLSAFSRYARCLAEHG 30

Db 141 LVTEYGSWRNRKLVFFSRYARTCFEAFDG 170
| | : | | : | | | | | | | | | |

RESULT 11

US-09-476-242-2

; Sequence 2, Application US/09476242

; Patent No. US2002014683A1

; GENERAL INFORMATION:

;; APPLICANT: BARNETT, Susan

;; APPLICANT: HARTOG, Karin

;; APPLICANT: MARTIN, Eric

;; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES

;; FILE REFERENCE: 1605.002

;; CURRENT APPLICATION NUMBER: US/09/476,242

;; CURRENT FILING DATE: 1999-12-30

;; NUMBER OF SEQ ID NOS: 26

;; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 2

;; LENGTH: 847

;; TYPE: PRT

;; ORGANISM: Human immunodeficiency virus

US-09-476-242-2

Query Match 25.8%; Score 49.5; DB 10; Length 847;

Best Local Similarity 33.3%; Pred. No. 96;

Matches 11; Conservative 5; Mismatches 10; Indels 7; Gaps 1;

QY 4 FWGDTLNCWM-----LSAFSRYARCLAEHGD 29
: | | : | | : | | | | | | | |

Db 786 YMGNLQYWIQELKNSAVSLFDAIAVAEGTD 818
: | | : | | : | | | | | | | |

RESULT 12

US-09-851-138-80

; Sequence 80, Application US/09851138

; Publication No. US20020183508A1

; GENERAL INFORMATION:

;; APPLICANT: MAERTENS, GEERT

;; STUYVER, LIEVEN

;; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS

;; NUMBER OF SEQUENCES: 207

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: ARNOLD, WHITE & DURKEE

;; STREET: P. O. BOX 4433

;; CITY: HOUSTON

;; STATE: TEXAS

;; COUNTRY: USA

;; ZIP: 77210-4433

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Microsoft Word 6.0 / ASCII text output

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/09/851,138

;; FILING DATE: 09-May-2001

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/836,075

;; FILING DATE: <Unknown>

;; APPLICATION NUMBER: EP 94870166.9

;; FILING DATE: 21 Oct 1994

;; APPLICATION NUMBER: EP 95870076.7

;; FILING DATE: 28 Jun 1995

;; ATTORNEY/AGENT INFORMATION:

;; NAME: KAMMERER, PATRICIA A.

;; REGISTRATION NUMBER: 29,775

;; REFERENCE/DOCKET NUMBER: INNS:004

;; INFORMATION FOR SEQ ID NO: 80:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 113 amino acids

;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 80:
US-09-851-138-80

Query Match 25.0%; Score 48; DB 9; Length 113;
Best Local Similarity 36.7%; Pred. No. 17;
Matches 11; Conservative 4; Mismatches 11; Indels 4; Gaps 1;

QY 3 TFWGDTLNCWMLSAFSAFYARCLAEHGHDGPT 32
Db 62 TSMGNTITCYV----KAAACRAAGIDAPT 87

RESULT 13

US-10-210-296-102
; Sequence 102, Application US/10210296
; Publication No. US20030021802A1
GENERAL INFORMATION:
; APPLICANT: Pfizer Products Inc.
; TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS, AND RELATED
; TITLE OF INVENTION: METHODS
; TITLE OF INVENTION: AND MATERIALS
; FILE REFERENCE: PC10589A
; CURRENT APPLICATION NUMBER: US/10/210,296
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/689,065
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 102
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-10-210-296-102

Query Match 25.0%; Score 48; DB 9; Length 526;
Best Local Similarity 40.0%; Pred. No. 93;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 8 TLNCWMLSAFSAFYARCLAEHGHDGPT 32
Db 158 TLNCSTITLLAKSFRSLPEKHNGLT 182

RESULT 14

US-10-016-283-34
; Sequence 34, Application US/10016283
; Patent No. US20020164702A1
GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: PCT/US96/20696
; PRIOR FILING DATE: 1996-12-13
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 34
; LENGTH: 1940
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-016-283-34

Query Match 25.0%; Score 48; DB 9; Length 1940;
Best Local Similarity 39.3%; Pred. No. 3.9e+02;
Matches 11; Conservative 1; Mismatches 14; Indels 2; Gaps 1;

QY 6 GDTLNL--CWMLSAFSAFYARCLAEHGHDGP 31

Db 328 GHTYNNICWRQAECCROQRAIPPKHQGP 355

RESULT 15

US-09-864-761-39145
; Sequence 39145, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aesomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39145
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004691.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.3
; OTHER INFORMATION: EST_HUMAN HIT: AW470102.1, EVALUATE 4.30e-01
; OTHER INFORMATION: SWISSPROT HIT: P01842, EVALUATE 1.90e+00
US-09-864-761-39145

Query Match 24.7%; Score 47.5; DB 10; Length 84;
Best Local Similarity 33.3%; Pred. No. 15;
Matches 10; Conservative 3; Mismatches 6; Indels 11; Gaps 2;

Qy 1 LGTFWGDTLNCWMLSAFSRYARCLAEHGDG 30
| | | | | : | | | |
Db 50 LDPFW-AISCW-----KCHVEPHDG 68

Search completed: May 29, 2003, 15:54:15
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 15:42:00 ; Search time 27 Seconds
(without alignments)
35.961 Million cell updates/sec

Title: US-09-989-481-4

-Perfect score: 192

Sequence: 1 LGTFWGTLCNWMLSAFSRVACLAEGHDGPTQ 33

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*

2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*

3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*

4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*

5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*

6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	100.0	33	4	US-08-928-862-4
2	51	26.6	465	4	US-08-914-375C-60
3	48	25.0	113	4	US-08-836-075A-80
4	48	25.0	1940	2	US-08-644-271-30
5	48	25.0	1940	4	US-09-077-955-34
6	47.5	24.7	865	4	US-09-612-204B-24
7	46.5	24.2	243	3	US-09-191-647-14
8	46.5	24.2	243	4	US-09-540-245A-14
9	46.5	24.2	243	4	US-09-540-153-14
10	46.5	24.2	771	4	US-09-188-930-183
11	46.5	24.2	1525	3	US-09-191-647-2
12	46.5	24.2	1525	4	US-09-540-245A-2
13	46.5	24.2	1525	4	US-09-540-153-2
14	45.5	23.7	196	4	US-09-129-030-28
15	45	23.4	2860	4	US-08-826-267-2
16	44.5	23.2	270	4	US-09-399-913-59
17	44.5	23.2	401	1	US-08-368-803-7
18	44.5	23.2	422	2	US-08-663-566A-5
19	44.5	23.2	422	2	US-08-023-610-5
20	44.5	23.2	422	2	US-08-288-065A-5
21	44.5	23.2	422	2	US-08-362-240A-5
22	44.5	23.2	422	4	US-08-804-372A-3
23	44.5	23.2	422	5	PCT-US95-10245-5
24	44	22.9	113	4	US-08-836-075A-78
25	44	22.9	222	6	5223425-6
26	44	22.9	238	4	US-08-944-483-44
27	44	22.9	238	6	5223425-5

28	44	22.9	250	6	5223425-4	Patent No. 5223425
29	44	22.9	253	6	5223425-8	Patent No. 5223425
30	44	22.9	425	1	US-08-190-802A-48	Sequence 48, Appl
31	44	22.9	425	4	US-08-477-346-48	Sequence 48, Appl
32	44	22.9	425	4	US-08-473-089-48	Sequence 48, Appl
33	44	22.9	425	4	US-08-487-072A-48	Sequence 107, App
34	43.5	22.7	178	3	US-08-906-769-107	Sequence 107, App
35	43.5	22.7	178	3	US-08-906-616-107	Sequence 107, App
36	43.5	22.7	178	4	US-08-817-795-107	Sequence 107, App
37	43.5	22.7	178	4	US-08-639-075A-107	Sequence 107, App
38	43.5	22.7	178	4	US-09-012-431-107	Sequence 107, App
39	43.5	22.7	178	4	US-09-012-692-107	Sequence 107, App
40	43.5	22.7	178	4	US-08-906-613-107	Sequence 107, App
41	43.5	22.7	178	5	PCT-US95-14442A-107	Sequence 107, App
42	43.5	22.7	225	4	US-09-004-731-13	Sequence 13, Appl
43	43.5	22.7	225	4	US-08-749-699-13	Sequence 13, Appl
44	43.5	22.7	225	4	US-09-004-729-13	Sequence 13, Appl
45	43.5	22.7	228	4	US-09-004-731-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-08-928-862-4

; Sequence 4, Application US/08928862

; Patent No. 6309877

; GENERAL INFORMATION:

; APPLICANT: Chau, Raymond M. W.

; TITLE OF INVENTION: Isolation and Use of Motoneuronotrophic Factors

; FILE REFERENCE: 12592-2

; CURRENT APPLICATION NUMBER: US/08/928.862

; CURRENT FILING DATE: 1997-09-12

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 33

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-928-862-4

Query Match 100.0%; Score 192; DB 4; Length 33;

Best Local Similarity 100.0%; Pred. No. 2e-20;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LGTFWGTLCNWMLSAFSRVACLAEGHDGPTQ 33

Db 1 LGTFWGTLCNWMLSAFSRVACLAEGHDGPTQ 33

RESULT 2

US-08-914-375C-60

; Sequence 60, Application US/08914375C

; Patent No. 6377893

; GENERAL INFORMATION:

; APPLICANT: Steven A. Benner

; NUMBER OF SEQUENCES: 74

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Steven A. Benner

; STREET: 1501 NW 68th Terrace

; CITY: Gainesville

; STATE: FL

; COUNTRY: United States

; ZIP: 32605-4147

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch diskette

; COMPUTER: Apple Macintosh

; OPERATING SYSTEM: MacIntosh 7.0

; SOFTWARE: Microsoft Word

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/914.375C

; FILING DATE: 19-Aug-1997

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

, FRAXION AFFILIATION NUMBER: 00/001,007

, FRISK AFFLICTION NUMBER: 00/001,001

; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 243
; TYPE: PRT
; ORGANISM: mouse
US-09-540-153-14

Query Match 24.2%; Score 46.5; DB 4; Length 243;
Best Local Similarity 50.0%; Pred. No. 49;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 11 CWMLSAFSRYARCLAEHGDG 30
| :||| :||| |
Db 99 CLPINFASYCKCL-EGHGG 117

RESULT 10

US-09-188-930-183
; Sequence 183, Application US/09188930A
; Patent No. 6150502

GENERAL INFORMATION:

; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011C1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 183
; LENGTH: 771
; TYPE: PRT
; ORGANISM: Rat
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (717)...(717)
US-09-188-930-183

Query Match 24.2%; Score 46.5; DB 4; Length 771;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 11 CWMLSAFSRYARCLAEHGDG 30
| :||| :||| |
Db 628 CLPINFASYCKCL-EGHGG 646

RESULT 11

US-09-191-647-2
; Sequence 2, Application US/09191647
; Patent No. 6046015

GENERAL INFORMATION:

; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/191,647
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: 60/065,544
; EARLIER FILING DATE: 1997-11-14
; EARLIER APPLICATION NUMBER: 60/081,057
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-191-647-2

Query Match 24.2%; Score 46.5; DB 3; Length 1525;
Best Local Similarity 50.0%; Pred. No. 3.9e+02;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 11 CWMLSAFSRYARCLAEHGDG 30
| :||| :||| |
Db 1381 CLPINFASYCKCL-EGHGG 1399

RESULT 12

US-09-540-245A-2
; Sequence 2, Application US/09540245A
; Patent No. 6270984

GENERAL INFORMATION:

; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-540-245A-2

Query Match 24.2%; Score 46.5; DB 4; Length 1525;
Best Local Similarity 50.0%; Pred. No. 3.9e+02;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 11 CWMLSAFSRYARCLAEHGDG 30
| :||| :||| |
Db 1381 CLPINFASYCKCL-EGHGG 1399

RESULT 13

US-09-540-153-2
; Sequence 2, Application US/09540153
; Patent No. 6270995

GENERAL INFORMATION:

; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,153
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/191,647
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: human
US-09-540-153-2

Query Match 24.2%; Score 46.5; DB 4; Length 1525;
Best Local Similarity 50.0%; Pred. No. 3.9e+02;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 11 CWMLSAFSRYARCLAEHGDG 30
Db 1381 CLPINFASYSCKCL-EGHGG 1399

RESULT 14

US-09-129-030-28
; Sequence 28, Application US/09129030A
; Patent No. 6242221
; GENERAL INFORMATION:
; APPLICANT: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION
; TITLE OF INVENTION: GENOMIC PPO CLONES
; FILE REFERENCE: 57072-PCT-US
; CURRENT APPLICATION NUMBER: US/09/129,030A
; CURRENT FILING DATE: 1998-08-04
; EARLIER APPLICATION NUMBER: AU PN7856
; EARLIER FILING DATE: 1996-02-05
; EARLIER APPLICATION NUMBER: AU P02361
; EARLIER FILING DATE: 1996-09-16
; EARLIER APPLICATION NUMBER: PCT/AU97/00041
; EARLIER FILING DATE: 1997-01-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28
; LENGTH: 196
; TYPE: PRT
; ORGANISM: APPLE
US-09-129-030-28

Query Match 23.7%; Score 45.5; DB 4; Length 196;
Best Local Similarity 40.7%; Pred. No. 53;
Matches 11; Conservative 2; Mismatches 9; Indels 5; Gaps 1;

QY 11 CWMLSAFSRY-----ARCLAEHGDGPT 32
Db 24 CWLFFPHRYLYFHERILAKLIDPT 50

RESULT 15

US-08-826-267-2
; Sequence 2, Application US/08826267
; Patent No. 5994070
; GENERAL INFORMATION:
; APPLICANT: Streuli, Michel
; TITLE OF INVENTION: No. 5994070e1 TRIO Molecules and Uses Related Thereto
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/826,267
; FILING DATE: 1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/014,214
; FILING DATE: 27 MARCH (1996)
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-010

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2860 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-826-267-2

Query Match 23.4%; Score 45; DB 2; Length 2860;
Best Local Similarity 39.1%; Pred. No. 1.3e+03;
Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
QY 6 GDTLNCWMLSAFSRYARCLAEHGH 28
Db 1711 GNTLRKWLTPVRLSSGRADGH 1733

Search completed: May 29, 2003, 15:46:26
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 15:33:30 ; Search time 69 Seconds
(without alignments)
63.729 Million cell updates/sec

Title: US-09-989-481-4
Perfect score: 192
Sequence: 1 LGTFWDTLNCWLSAFSRVACRLAEGHDGPTQ 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	100.0	33	AAW59046	Human MNTF1-F6 pro
2	56.5	29.4	584	AAW90949	Comamonas testoste
3	55.5	28.9	227	AAU19525	Human diagnostic a
4	55	28.6	694	AAW93813	Herbicidally activ
5	53.5	27.9	326	AAW23939	Hepatitis B virus
6	53.5	27.9	384	AAW94802	Human protein sequ
7	53	27.6	794	AAW25667	Novel human diagno
8	52	27.1	563	AAW10330	Gene product with
9	51	26.6	56	ABG60253	Human ovarian anti
10	51	26.6	56	AAW94423	Human reproductive

11	51	26.6	56	23	ABG61724	Novel ovarian rela
12	51	26.6	466	23	AAE23630	Escherichia coli 6
13	51	26.6	474	22	AAU34677	E. coli cellular p
14	51	26.6	1207	22	AAW78524	Human protein SEQ
15	51	26.6	1207	22	AAW78524	Amino acid sequenc
16	51	26.6	1222	22	ABW4604	Human precursor pr
17	51	26.6	1222	22	AAW79508	Human protein SEQ
18	51	26.6	1225	22	ABG24444	Novel human diagno
19	51	26.6	1258	22	ABG24819	Novel human diagno
20	50	26.0	726	22	ABG68590	ACNGC2/DND1 prote
21	50	26.0	726	23	ABW93489	Herbicidally activ
22	50	26.0	844	22	ABW61902	Drosophila melanog
23	49.5	25.8	842	23	ABW60211	HIV Env isolate SF
24	49.5	25.8	847	21	AAW97073	Variant HIV-1 SF16
25	49.5	25.8	1054	22	ABW60410	Drosophila melanog
26	49	25.5	414	23	ABW54212	Lactococcus lactis
27	49	25.5	665	19	AAW54425	Human PS12 protei
28	49	25.5	665	21	AAW08415	Protein encoded by
29	48.5	25.3	150	22	AAU87494	Novel central nerv
30	48	25.0	94	22	ABG19025	Novel human diagno
31	48	25.0	113	17	AAW96565	Hepatitis C virus
32	48	25.0	116	22	ABG30144	Novel human diagno
33	48	25.0	526	22	AAW78608	Lawsonia intracell
34	48	25.0	731	22	ABG30155	Novel human diagno
35	48	25.0	743	23	ABW92349	Herbicidally activ
36	48	25.0	839	23	AAE23384	Human intracellular
37	47.5	24.7	84	22	ABW38826	Peptide #6332 enco
38	47.5	24.7	84	22	ABW23847	Protein #5846 enco
39	47.5	24.7	84	22	AAW59470	Human bone marrow
40	47.5	24.7	84	22	AAW72034	Peptide #5847 enco
41	47.5	24.7	84	22	AAW19413	Peptide #6337 enco
42	47.5	24.7	84	22	AAW32300	Human peptide enco
43	47.5	24.7	84	23	ABG41849	Herbicidally activ
44	47.5	24.7	273	23	ABW91902	Arabidopsis thalia
45	47.5	24.7	433	21	AAW50117	

ALIGNMENTS

RESULT 1

AAW59046
ID AAW59046 standard; Protein; 33 AA.
AC AAW59046;
DT 11-AUG-1998 (first entry)
DE Human MNTF1-F6 protein fragment.
KW Motoneuronotrophic factor; MNTF-1; MNTF1-F6; human; axon regeneration;
KW motoneuron; diagnosis; treatment; disease; wound healing; scar tissue;
KW keloid.

XX Homo sapiens.

XX OS

XX PN WO9813492-A2.

XX PD 02-APR-1998.

XX PF 22-SEP-1997; 97WO-US17142.

XX PR 12-SEP-1997; 97US-0928862.

XX PR 27-SEP-1996; 96US-0026792.

XX PR 15-NOV-1996; 96US-0751225.

XX PA (KMBI-) KM BIOTECH INC.

XX PI Chau RMW;

XX WPI; 1998-230705/20.

XX N-PSDB; AAW1174E.

XX

XX PS Claim 27; Page 464; 522pp; English.

XX CC The invention relates to polynucleotides (I) encoding diagnostic and

XX CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,

XX CC and proteins involved in growth and development and receptors. (I) and

XX CC (II) may be used in the prevention, diagnosis and treatment of diseases

XX CC associated with inappropriate DITHP expression. For example, (I) and

XX CC (II) may be used to treat disorders associated with decreased polypeptide

XX CC expression by rectifying mutations or deletions in a patient's genome,

XX CC that affect the activity of the DITHPs, by expressing inactive proteins

XX CC or supplementing the patient's own production of them. (I) and (II)

XX CC may be used to treat diseases, for example, cell proliferative disorder,

XX CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,

XX CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,

XX CC (I) may be used to produce the DITHPs, by inserting the nucleic acids

XX CC into a host cell and culturing the cell to express the protein. (I) and

XX CC its complementary sequences may also be used as DNA probes in diagnostic

XX CC assays to detect and quantitate the presence of similar nucleic acids in

XX CC samples, and therefore which patients may be in need of restorative

XX CC therapy. (II) may also be used as antigens in the production of

XX CC antibodies against DITHPs and in assays to identify modulators of DITHP

XX CC expression and activity. The anti-DITHP antibodies and antagonists may

XX CC also be used to down regulate expression and activity. The anti-DITHP

XX CC antibodies may also be used as diagnostic agents for detecting the

XX CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbent

XX CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and

XX CC therapeutic (DITHP) polypeptides of the invention.

XX SQ Sequence 227 AA;

Query Match 28.9%; Score 55.5; DB 22; Length 227;
Best Local Similarity 22.6%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 8; Indels 35; Gaps 2;

Qy 4 FWGDTLNCW-----MLSAFSRY-----ARCLAEHG 28
Db 125 FWGQRCNCGSRASAPLFSAFSEFPAGGVFSFDTGFRSGLSGGLSFCMSYGS 184
Qy 29 DG 30
Db 185 DG 186

RESULT 4
ABB93813
ID ABB93813 standard; Protein; 694 AA.

XX AC ABB93813;
XX DT 31-MAY-2002 (first entry)
XX DE Herbicidally active polypeptide SEQ ID NO 3024.
XX KW Herbicidal; plant; agriculture; herbicide.
XX OS Arabidopsis thaliana.
XX PN WO200210210-A2.
XX PD 07-FEB-2002.
XX PF 28-AUG-2001; 2001WO-EP09892.
XX PR 28-AUG-2001; 2001WO-EP09892.
XX PA (FARB) BAYER AG.

XX PI Tietjen K, Weidner M;
XX KW WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,

PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
XX Claim 5; SEQ ID NO 3024; 261pp + Sequence Listing; English.
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.

XX SQ Sequence 694 AA;

Query Match 28.6%; Score 55; DB 23; Length 694;
Best Local Similarity 35.9%; Pred. No. 67;
Matches 14; Conservative 2; Mismatches 9; Indels 14; Gaps 2;

Qy 2 GT-FWGYTLN-----CWMLSAFSRYARCLAE 26
Db 255 GTVWVGTAALNMIAFYFAHAAGACWYLLGVQVSARCLAE 293

RESULT 5
AAB23939
ID AAB23939 standard; Protein; 326 AA.

XX AC AAB23939;
XX DT 18-JAN-2001 (first entry)

XX DE Hepatitis B virus protein bound arrestin protein sequence SEQ ID NO:2.
XX KW Hepatitis B virus; HBV; arrestin; binding.

XX OS Hepatitis B virus.
XX PN CN1257919-A.

XX PD 28-JUN-2000.

XX PF 21-DEC-1998; 98CN-0125693.

XX PR 21-DEC-1998; 98CN-0125693.

XX PA (UYFU-) UNIV FUTAN.

XX PI Yu L, Wang X, Fu Q;

XX DR WPI; 2000-544292/50.

XX DR N-PSDB; AAA99087.

XX PT Hepatitis B virus protein bound arrestin -

XX PS Claim 1; Page 13; 16pp; Chinese.

XX CC The present sequence represents a specifically claimed protein
XX sequence from the present invention. The present invention describes
XX Hepatitis B virus (HBV) protein bound arrestin. Also described is a
XX method for the preparation of the novel protein and polynucleotide of
XX the invention.

XX SQ Sequence 326 AA;

Query Match 27.9%; Score 53.5; DB 21; Length 326;
Best Local Similarity 40.0%; Pred. No. 47;
Matches 12; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Qy 4 FWGDTLNCWMLSAFSRYARCLAEHGDPQTQ 33

Db 739 DALNLFPLQINPHFTNALAECHKGETR 765

RESULT 8
AAR10330
ID AAR10330 standard; Protein: 563 AA.
XX AC AAR10330;
XX DT 05-APR-1991 (first entry)
XX DE Gene product with lipase activity.
XX KW ATCC 34614.
XX OS Geotrichum candidum.
XX PN JP02299588-A.
XX PD 11-DEC-1990.
XX PF 27-MAR-1989; 89JP-0074721.
XX PR 27-MAR-1989; 89JP-0074721.
XX PA (KURK) KURITA WATER IND KK.
XX PA (OSAKA) OSAKA CITY.
XX DR WPI: 1991-027567/04.
XX DR N-PSDB; AAQ10313.
XX

Gene for coding protein with lipase activity - is prep'd. from messenger ribonucleic acid of geo-trichum candidum ATCC 34614
Claim 1; Fig 4; 12pp; Japanese.
The gene product may be isolated from a transformed expression system, and may be enhanced with stability in heat, alkalai, acid and organic solvent by position-specific modulation.
Sequence 563 AA;
Query Match 27.1%; Score 52; DB 12; Length 563;
Best Local Similarity 40.0%; Pred. No. 1.3e+02;
Matches 14; Conservative 1; Mismatches 14; Indels 6; Gaps 1;
Y 1 LGTFWGDTL-----NCWMLSAFSRYARCLAEHGD 29
||||| | | | | | | | | | | | | | | | | |
478 LGTFHGSDDLFLQYYAGPWSSAYRRYFISFANHHDD 512

Db

RESULT 9
ABG60253
ID ABG60253 standard; Protein: 56 AA.
XX AC ABG60253;
XX DT 13-AUG-2002 (first entry)
XX DE Human ovarian antigen #15.
XX

Human; ovarian antigen; ovary disorder; breast disorder;
neoplastic disorder; cancer; infectious disease; inflammatory disease;
reproductive system disorder; autoimmune disorder; Alzheimer's disease;
blood-related disorder; hyperproliferative disorder; hair loss;
urinary system disorder; cardiovascular disorder; arrhythmia;
respiratory disorder; musculoskeletal system disorder;
neural activity disorder; neurological disorder; endocrine disorder;
gastrointestinal disorder; liver disorder; pancreatic disorder;
gall bladder disorder; large intestine disorder; developmental disorder;
inherited disorder; wound healing; skin aging; food additive;
preservative.

XX

OS Homo sapiens.
XX WO200155329-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01360.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 14-SEP-2000; 2000US-0232398.
XX PR 17-NOV-2000; 2000US-0249300.
XX PR 01-DEC-2000; 2000US-0250160.
XX PR 08-DEC-2000; 2000US-0251868.
XX PR 08-DEC-2000; 2000US-0251990.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-476195/51.
XX N-PSDB; ABK72056.
XX

Novel isolated human ovarian related polypeptide useful for diagnosis/treatment of disorders of ovary and breast such as neoplastic disorders, infectious diseases, inflammatory diseases, and reproductive disorders
Claim 11; SEQ ID NO 83; 524pp; English.
The invention relates to isolated ovarian related polypeptide (ovarian antigen) comprising a sequence at least 90% identical to a sequence selected from a polypeptide fragment, domain, epitope or full-length protein of a sequence (S1) appearing as ABG60253-ABG60296 having biological activity, or a variant, allelic variant or species homologue of S1. Also included are the cDNA clones encoding the proteins of S1. S1, an anti-S1 antibody and the cDNA are useful for diagnosing, preventing, treating or ameliorating a medical condition in mammalian subject especially diseases and/or disorders of the ovary and/or breast such as neoplastic disorders (such as ovarian Krukenberg tumour and cancer), infectious diseases (e.g., mastitis, oophoritis), inflammatory diseases (e.g., abscesses), reproductive system disorders (Paget's disease), autoimmune disorders (systemic lupus erythematosus, rheumatoid arthritis), blood-related disorders (sickle cell anaemia), hyperproliferative disorders, urinary system disorders (glomerulonephritis), cardiovascular disorders (arrhythmias), respiratory disorders, musculoskeletal system disorders, neural activity and neurological disorders (Alzheimer's disease and Parkinson's disease), endocrine disorders (Addison's disease), gastrointestinal disorders (inflammatory disorders), liver disorders (biliary liver cirrhosis), pancreatic and gall bladder disorders, disorders of the large intestine, developmental and inherited disorders, diseases at the cellular level, and wound healing and epithelial cell proliferation. They are also useful to prevent skin aging, for preventing hair loss, to maintain organs before transplantation or for supporting cell culture of primary tissues, to modulate mammalian characteristics such as body height, to modulate mammalian metabolism, to change a mammal's mental or physical state, and as food additive or preservative. The present sequence represents an ovarian antigen, S1 protein of the invention.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
Sequence 56 AA;
Query Match 26.6%; Score 51; DB 22; Length 56;

PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 06-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465570/50.

N-PSDB; AAL00393.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition -

Claim 11; SEQ ID NO 3081; 1297pp + Sequence Listing; English.

The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a protein of the invention.

Sequence 56 AA;

Query Match 26.6%; Score 51; DB 22; Length 56;

Best Local Similarity 57.9%; Pred. No. 15;

Matches 11; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

Oy 9 LNCWMLSAFSRYAR--CLA 25

||||| | | | |

25 LNCWHLSCFNHALRLSCLA 43

RESULT 11

ABG61724

ID ABG61724 standard; Protein: 56 AA.

XX AC ABG61724;

DT 26-AUG-2002 (first entry)

XX DE

XX Novel ovarian related polypeptide #15.

XX Ovarian related polypeptide; neoplastic disorder; tumour; ovarian cancer;
 KW hyperproliferative disorder; adult acute lymphocytic leukaemia;
 KW breast cancer; reproductive system disorder; tuberculosis; arthritis;
 KW immune system disorder; Chediak-Higashi's syndrome; neonatal neutropenia;
 KW autoimmune disorder; Hashimoto's thyroiditis; inflammatory disorder;
 KW septic shock; multiple sclerosis; central nervous system disorder;
 KW neurological disorder; allergy; Parkinson's disease; Alzheimer's disease;
 KW cardiovascular disorder; atherosclerosis; blood related disorder;
 KW respiratory disorder; urinary system disorder; musculoskeletal disorder;
 KW osteoporosis; wound healing; endocrine disorder; infectious disease;
 KW gastrointestinal disorder; transplantation; food additive; preservative.

OS Homo sapiens.

XX US2002045230-A1.

PN

XX 18-APR-2002.
 PD 20-JUL-2001; 2001US-0908711.
 XX 31-JAN-2000; 2000US-179065P.
 PR 04-FEB-2000; 2000US-180628P.
 PR 24-FEB-2000; 2000US-184664P.
 PR 02-MAR-2000; 2000US-186350P.
 PR 16-MAR-2000; 2000US-189874P.
 PR 17-MAR-2000; 2000US-190076P.
 PR 18-APR-2000; 2000US-198123P.
 PR 19-MAY-2000; 2000US-205515P.
 PR 07-JUN-2000; 2000US-209467P.
 PR 28-JUN-2000; 2000US-214886P.
 PR 30-JUN-2000; 2000US-215135P.
 PR 07-JUL-2000; 2000US-216647P.
 PR 07-JUL-2000; 2000US-216880P.
 PR 11-JUL-2000; 2000US-217487P.
 PR 11-JUL-2000; 2000US-217496P.
 PR 14-JUL-2000; 2000US-218290P.
 PR 26-JUL-2000; 2000US-220963P.
 PR 26-JUL-2000; 2000US-220964P.
 PR 14-AUG-2000; 2000US-224518P.
 PR 14-AUG-2000; 2000US-224519P.
 PR 14-AUG-2000; 2000US-225213P.
 PR 14-AUG-2000; 2000US-225214P.
 PR 14-AUG-2000; 2000US-225266P.
 PR 14-AUG-2000; 2000US-225267P.
 PR 14-AUG-2000; 2000US-225268P.
 PR 14-AUG-2000; 2000US-225270P.
 PR 14-AUG-2000; 2000US-225447P.
 PR 14-AUG-2000; 2000US-225757P.
 PR 14-AUG-2000; 2000US-225758P.
 PR 14-AUG-2000; 2000US-225759P.
 PR 18-AUG-2000; 2000US-226279P.
 PR 22-AUG-2000; 2000US-226681P.
 PR 22-AUG-2000; 2000US-226868P.
 PR 22-AUG-2000; 2000US-227182P.
 PR 23-AUG-2000; 2000US-227009P.
 PR 30-AUG-2000; 2000US-228924P.
 PR 01-SEP-2000; 2000US-229287P.
 PR 01-SEP-2000; 2000US-229343P.
 PR 01-SEP-2000; 2000US-229344P.
 PR 01-SEP-2000; 2000US-229345P.
 PR 05-SEP-2000; 2000US-229509P.
 PR 05-SEP-2000; 2000US-229513P.
 PR 06-SEP-2000; 2000US-230437P.
 PR 06-SEP-2000; 2000US-230438P.
 PR 08-SEP-2000; 2000US-231242P.
 PR 08-SEP-2000; 2000US-231243P.
 PR 08-SEP-2000; 2000US-231244P.
 PR 08-SEP-2000; 2000US-231413P.
 PR 08-SEP-2000; 2000US-231414P.
 PR 08-SEP-2000; 2000US-232080P.
 PR 08-SEP-2000; 2000US-232081P.
 PR 12-SEP-2000; 2000US-231968P.
 PR 14-SEP-2000; 2000US-232397P.
 PR 14-SEP-2000; 2000US-232398P.
 PR 14-SEP-2000; 2000US-232399P.
 PR 14-SEP-2000; 2000US-232400P.
 PR 14-SEP-2000; 2000US-232401P.
 PR 14-SEP-2000; 2000US-233063P.
 PR 14-SEP-2000; 2000US-233064P.
 PR 14-SEP-2000; 2000US-233065P.
 PR 21-SEP-2000; 2000US-234223P.
 PR 21-SEP-2000; 2000US-234274P.
 PR 25-SEP-2000; 2000US-234997P.
 PR 25-SEP-2000; 2000US-234998P.
 PR 26-SEP-2000; 2000US-235484P.
 PR 27-SEP-2000; 2000US-235834P.
 PR 27-SEP-2000; 2000US-235836P.
 PR 29-SEP-2000; 2000US-236327P.

PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254057P.
PR 05-JAN-2001; 2001US-259678P.
PR 17-JAN-2001; 2001US-0764853.
PR 17-JAN-2001; 2001US-0764856.
PR 17-JAN-2001; 2001US-0764864.
PR 17-JAN-2001; 2001US-0764867.
PR 17-JAN-2001; 2001US-0764868.

PR 17-JAN-2001; 2001US-0764869.
PR 17-JAN-2001; 2001US-0764870.
PR 17-JAN-2001; 2001US-0764874.
PR 17-JAN-2001; 2001US-0764882.
PR 17-JAN-2001; 2001US-0764888.
PR 17-JAN-2001; 2001US-0764891.
PR 17-JAN-2001; 2001US-0764892.
PR 17-JAN-2001; 2001US-0764896.
PR 17-JAN-2001; 2001US-0764898.
PR 17-JAN-2001; 2001US-0764902.
PR 17-JAN-2001; 2001US-0764905.
PR 17-JAN-2001; 2001WO-US01239.
PR 17-JAN-2001; 2001WO-US01307.
PR 17-JAN-2001; 2001WO-US01312.
PR 17-JAN-2001; 2001WO-US01320.
PR 17-JAN-2001; 2001WO-US01329.
PR 17-JAN-2001; 2001WO-US01334.
PR 17-JAN-2001; 2001WO-US01336.
PR 17-JAN-2001; 2001WO-US01339.
PR 17-JAN-2001; 2001WO-US01340.
PR 17-JAN-2001; 2001WO-US01341.
PR 17-JAN-2001; 2001WO-US01344.
PR 17-JAN-2001; 2001WO-US01345.
PR 17-JAN-2001; 2001WO-US01347.
PR 17-JAN-2001; 2001WO-US01348.
PR 17-JAN-2001; 2001WO-US01360.
XX
PA (ROSE/) ROSEN C A.
PA (RUBI/) RUBIN S M.
PA (BARA/) BARASH S C.
XX

Query Match 26.6%; Score 51; DB 23; Length 56;
Best Local Similarity 57.9%; Pred. No. 15;
Matches 11; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

QY 9 LNCWMLSAFSRYAR--CLA 25
|||||:|
Db 25 LNCWHLSCFNHALRLSCLA 43

RESULT 12

AAE23630
ID AAE23630 standard; Protein; 466 AA.

XX

AC AAE23630;

XX

DT 27-AUG-2002 (first entry)

XX

DE Escherichia coli 6-phospho-strand-glucosidase #1.

XX Secondary structural element; genome-sized database; EC 3.2.1.86;
KW enzyme; 6-phospho-strand-glucosidase.

XX

OS Escherichia coli.

XX

FH Key Location/Qualifiers

FT Region 9..11

FT /note= "Strand A"

FT Region 12..103

FT /note= "CG dipeptide parse"

FT Region 14..20

FT /note= "Strand B"

FT Region 48..59

FT /note= "Parse region"

FT Region 49..51

FT /note= "Strand A"

FT Region 50..54

FT /note= "Parsing string (PGDSG)"

FT Region 53..60

FT /note= "Strand B"

FT Region 62..63

FT /note= "CG dipeptide parse"

```

FT Region 71..81 /note= "Parse region"
FT Region 72..84 /note= "Helix 1"
FT Region 73..78 /note= "Beta strand"
FT Region 89..93 /note= "Strand C"
FT Region 95..107 /note= "Helix 1"
FT Active-site 95..102 /note= "Active site a"
FT Region 111..115 /note= "Strand C"
FT Region 116..130 /note= "Helix 2"
FT Active-site 117..125 /note= "Active site a"
FT Region 136..140 /note= "Strand D"
FT Region 138..153 /note= "Helix 2"
FT Active-site 141 /note= "Active site b"
FT Region 158..177 /note= "Helix 3"
FT Region 159..163 /note= "Strand D"
FT Active-site 164..166 /note= "Active site b"
FT Region 181..198 /note= "Helix 3"
FT Region 182..185 /note= "Strand E"
FT Active-site 184..187 /note= "Active site c"
FT Active-site 205..209 /note= "Active site c"
FT Region 205..208 /note= "Strand E"
FT Active-site 207..209 /note= "Active site c"
FT Region 212..215 /note= "Beta strand"
FT Region 212..226 /note= "Helix 4"
FT Region 219..221 /note= "Beta strand"
FT Region 242..245 /note= "Strand X"
FT Region 248..268 /note= "Helix 4"
FT Region 259..273 /note= "Helix X"
FT Region 275..280 /note= "Strand Y"
FT Region 286..293 /note= "Helix 5"
FT Region 318..320 /note= "Strand Y"
FT Region 323..327 /note= "Strand F"
FT Region 329..339 /note= "Helix Y"
FT Region 332..342 /note= "Helix 5"
FT Region 375..382 /note= "Strand Z"
FT Region 381..388 /note= "Strand F"
FT Region 385..398 /note= "Helix 6"
FT Region 404..407

FT Active-site 408..410 /note= "Strand G"
FT Region 431..448 /note= "Active site d"
FT Region 446..452 /note= "Helix 7"
FT Region 450..454 /note= "Strand 2"
FT Region 456..459 /note= "Strand H"
FT Region 456..469 /note= "Strand I"
FT Region 460..462 /note= "Helix 6"
FT Region 464..467 /note= "Dipeptide GP parse"
FT Region 476..479 /note= "Strand J"
FT Region 478..482 /note= "Strand G"
FT Region 480..482 /note= "Strand K"
FT Active-site 480..482 /note= "Active site d"
FT Region 496..509 /note= "Helix 8"
FT Region 497..517 /note= "Helix 7"
FT Region 521..525 /note= "Strand H"
FT Region 527..530 /note= "Strand I"
FT Region 535..539 /note= "Strand J"
FT Region 548..554 /note= "Strand K"
FT Region 563..576 /note= "Helix 8"

US6377893-B1.
23-APR-2002.
19-AUG-1997; 97US-0914375.
25-MAR-1992; 92US-0857224.
(BENNY) BENNER S A.
Benner SA;
WPI; 2002-424771/45.
Methods for excluding or detecting homology between protein families,
useful e.g. for identifying in vitro properties of proteins important
for physiological activity.
Example 5; Column 147-150; 99pp; English.
The invention relates to a method for excluding homology between
two protein families. The method involves constructing models for
secondary structural elements for each family; aligning secondary
structural elements of one family with the secondary structural
elements from the other family around sequence motifs; determining
whether secondary structural elements flanking the sequence motifs
in one family are congruent to secondary structural elements in
the other family; so as to determine if the families are related
by common ancestry or not. The method is used to confirm/deny the
hypothesis that proteins are homologous and related methods are
used to identify mutations during divergent evolution of proteins,
to identify in vitro properties of proteins that are important for
physiological activity and to generate genome-sized databases.
The present sequence is Escherichia coli 6-phospho-strand-glucosidase
(BC 3.2.1.86). This sequence is used in the exemplification of the

```

CC invention.
XX
SQ Sequence 466 AA;
Query Match 26.6%; Score 51; DB 23; Length 466;
Best Local Similarity 43.3%; Pred. No. 1.5e+02;
Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;
QY 1 LGTFWGDYLNWMLSAFSAFYARCLAEHGDG 30
| | : | | : : | | | | | | | | | |
DB 138 LVTEYGSWRNRKLVFEFFSRYARTCFEAFDG 167

RESULT 13
AAU34677
ID AAU34677 standard; Protein; 474 AA.
XX
AC AAU34677;
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #258.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Travick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR N-PSDB; AAS52536.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10270; 511pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 474 AA;
Query Match 26.6%; Score 51; DB 22; Length 474;
Best Local Similarity 43.3%; Pred. No. 1.5e+02;
Matches 13; Conservative 3; Mismatches 14; Indels 0; Gaps 0;
QY 1 LGTFWGDYLNWMLSAFSAFYARCLAEHGDG 30
| | : | | : : | | | | | | | | | |
DB 141 LVTEYGSWRNRKLVFEFFSRYARTCFEAFDG 170

RESULT 14
AAU78524
ID AAU78524 standard; Protein; 1207 AA.
XX
AC AAU78524;
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1186.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK51657.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 3434-3436; 6221pp; English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAU78323-AAU80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAU80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

XX SQ Sequence 1207 AA;

Query Match 26.6%; Score 51; DB 22; Length 1207;
Best Local Similarity 56.2%; Pred. No. 4.1e+02;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 18 SRYARCLAEHGHDGPTQ 33
| | | | | : | | | | |
Db 841 SMYARCISEGEDATCQ 856

RESULT 15

AAB84604
ID AAB84604 standard; Protein; 1207 AA.

XX AC AAB84604;

XX 05-SEP-2001 (first entry)

XX Amino acid sequence of endothelial growth factor.

XX Growth factor; protein inhibitor; protease; damaged tissue;
XX platelet-derived growth factor; PDGF; fibroblast growth factor; FGF;
XX connective tissue derived growth factor; CTGF; chrysalin; VEGF;
XX keratinocyte-derived growth factor; KGF; epidermal growth factor; EGF;
XX transforming growth factor-beta; TGF-beta; matrix metalloproteinase; MMP;
XX granulocyte macrophage colony stimulating factor; GM-CSF; uPA;
XX vascular endothelial growth factor; urokinase plasminogen activator;
XX dermal ulcer; wound.

XX Homo sapiens.

XX WO200149309-A2.

XX 12-JUL-2001.

XX 21-DEC-2000; 2000WO-IB01935.

XX 29-DEC-1999; 99GB-0030768.

XX (PFI2) PFIZER LTD.

XX (PFI2) PFIZER INC.

XX Davies MJ, Huggins JP, McIntosh FS, Occleston NL;

XX WPI; 2001-418351/44.

XX N-PSDB; AAH28219.

XX Composition for the treatment of damaged tissue i.e. chronic wounds and
XX dermal ulcers comprises an inhibitor agent i.e. a protease and a growth
XX factor .

XX Disclosure; Page 549; 572pp; English.

XX The specification describes a pharmaceutical composition, comprising
XX a growth factor, an inhibitor agent, i.e. a protease. The inhibitor
XX agent inhibits the action of at least one specific adverse protein,
XX i.e. a protease, that is upregulated in a damaged tissue such as a
XX wound environment. Growth factors which are included in the composition
XX of the invention are platelet-derived growth factor (PDGF), fibroblast
XX growth factor (FGF), connective tissue derived growth factor (CTGF),
XX keratinocyte-derived growth factor (KGF), transforming growth
XX factor-beta (TGF-beta), granulocyte macrophage colony stimulating factor
XX (GM-CSF), epidermal growth factor (EGF), vascular endothelial growth
XX factor (VEGF), and chrysalin. Inhibitors which are included in the
XX composition of the invention include inhibitors of urokinase-type
XX plasminogen activator (uPA) and matrix metalloproteinase (MMP). The
XX composition is useful for the treatment of chronic damaged tissue, i.e.
XX wounds and dermal ulcers. The present sequence represents a human EGF,
XX and is used to produce the composition of the invention.

XX Sequence 1207 AA;

Query Match 26.6%; Score 51; DB 22; Length 1207;
Best Local Similarity 56.2%; Pred. No. 4.1e+02;

Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 18 SRYARCLAEHGHDGPTQ 33
| | | | | : | | | | |
Db 841 SMYARCISEGEDATCQ 856

Search completed: May 29, 2003, 15:43:06
Job time : 71 secs

THIS PAGE BLANK (USPTO)